

SkyIntelligent – 工业大数据科学云平台实践

南京天数信息科技有限公司 – 倪岭

- 动机 背景
- 大数 平台演进
- 设计理念
- 系统
- 底层 件 计算引 性能
- 中间层 容器平台 象
- 顶层 前 交 便利
- 工作流程
- 展望





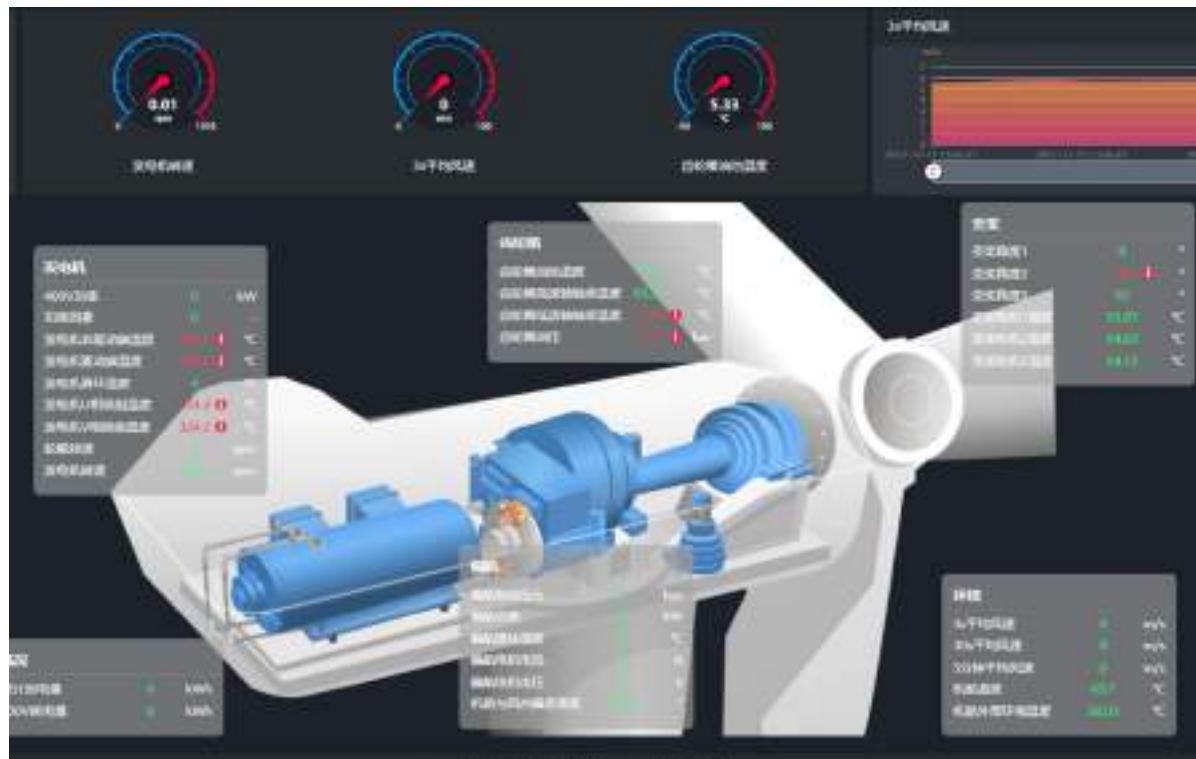
生产过程改进工艺改良



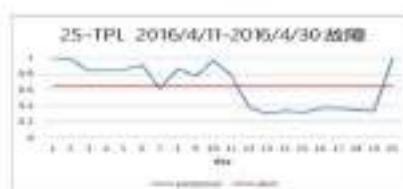
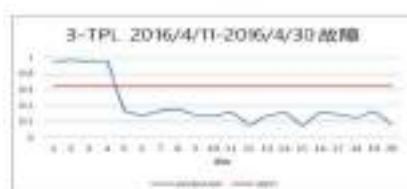
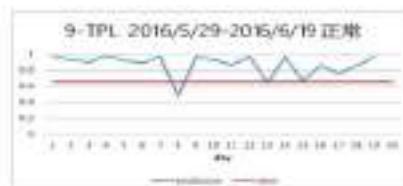
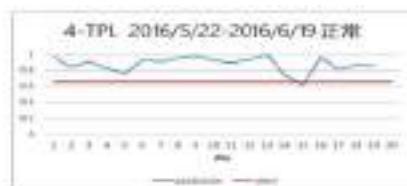
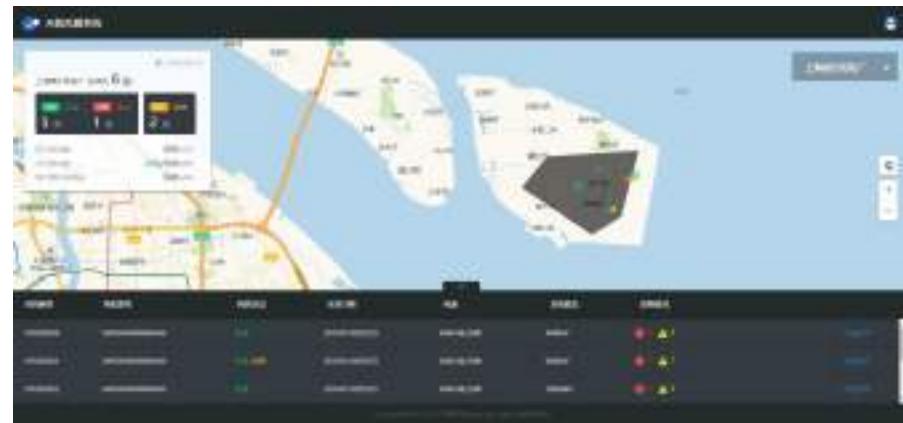
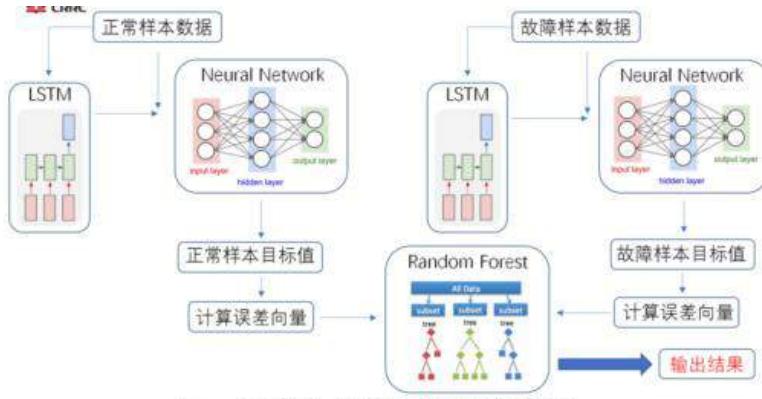
维护成本降低



更科学、高效的决策与管理



成本比较	之前	现在
齿轮箱故障	200万	10万
叶根螺栓断裂	120万	5万
叶片航拍视频分析 (1000台风机)	100万	5万

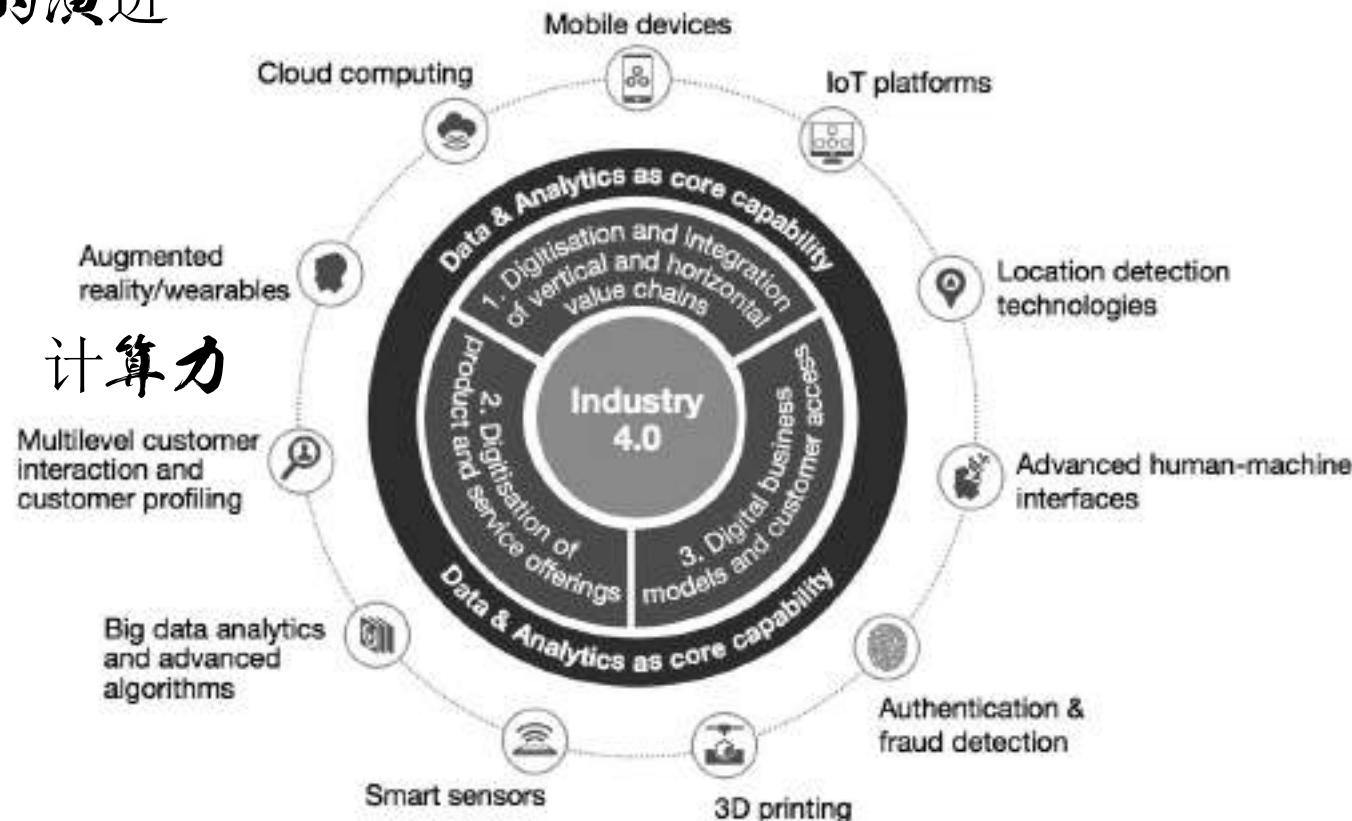


• 数 分 求的演进

• 数 算法 计算力

• 企业私有云

Industry 4.0 framework and contributing digital technologies



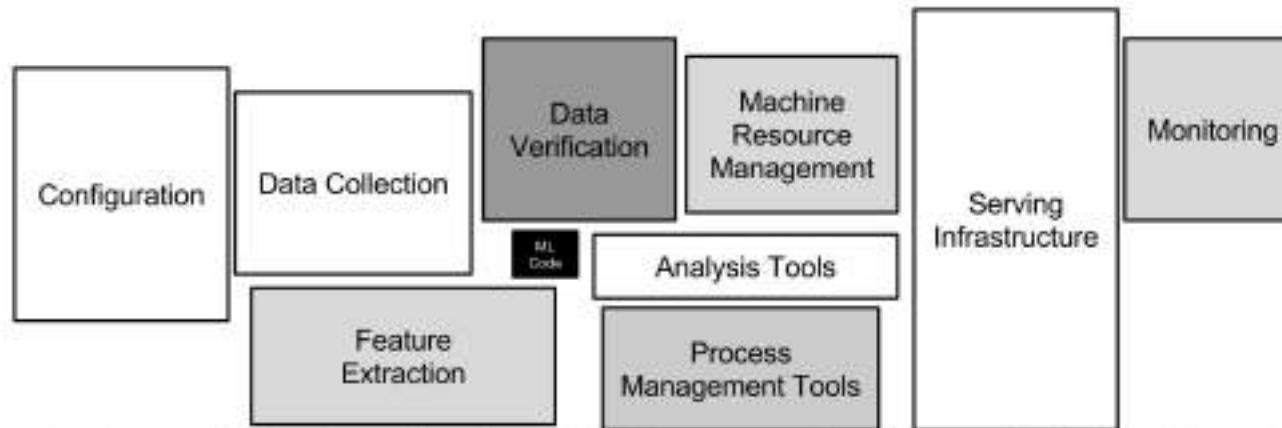


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

- 机器学习 深度学习 不 一 求
- 大数 平台 机器学习平台 深度学习平台 服务平
台
- 多平台
- 统一平台
- 业级私有数 科学 平台

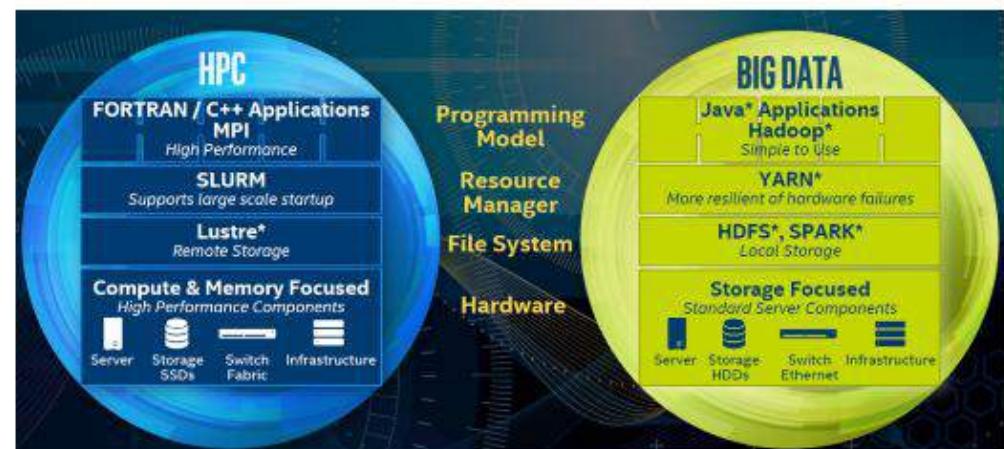
深度学习 化

— 高速 联
— 加速器

— 快速存储

合

THE CONVERGENCE OF HPC AND BIG DATA ANALYTIC SOLUTIONS



Other brands and names are the property of their respective owners.

Sources: Daniel Reed and Jack Dongarra, Exascale Computing and Big Data in Communications of the ACM Journal, July 2015 (Vol 58, No.7), and Intel analysis.

- 对接现有大数 平台 软 件资源重用
- 资源 拟化 弹性化
- 一 式 全流程 模块化
- 高性能 高可用
- 多 户 合作 共
- 易扩展 可用户自定义

在线
开发
调试

数
读取
转换

特
工程
存取

可视化
数
模型

模型
调试
评

任务
提交
执行

在线
部署
预测

机器学习

数

流处理

图计算

全文

数
存取

部署

维

监

管理

户

容器 管理

资源
管理

镜像
管理

服务
编

弹性
缩

负载
均

服务
发现

健康
检查

灰度
发布

故障
容错

权限
制

日志
集

监
告警

多
户
资源

生态 组件

高性能分布式计算引

深度学习

等

等

计算资源

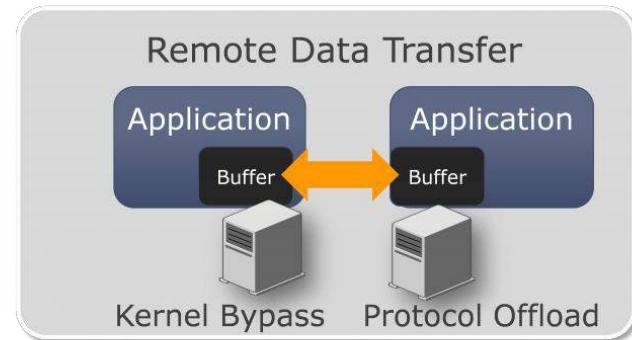
他 件资源

• 远程直接数 存取 的全面使用

• 动态编译技术

优化

• 混合存储优化



Low Latency, High Performance Data Transfers



InfiniBand - 56Gb/s

RoCE* - 40Gb/s

* RDMA over Converged Ethernet

- 深度学习

—

- 数库作为分中间件 前交式分

—

- 优化的优化

—

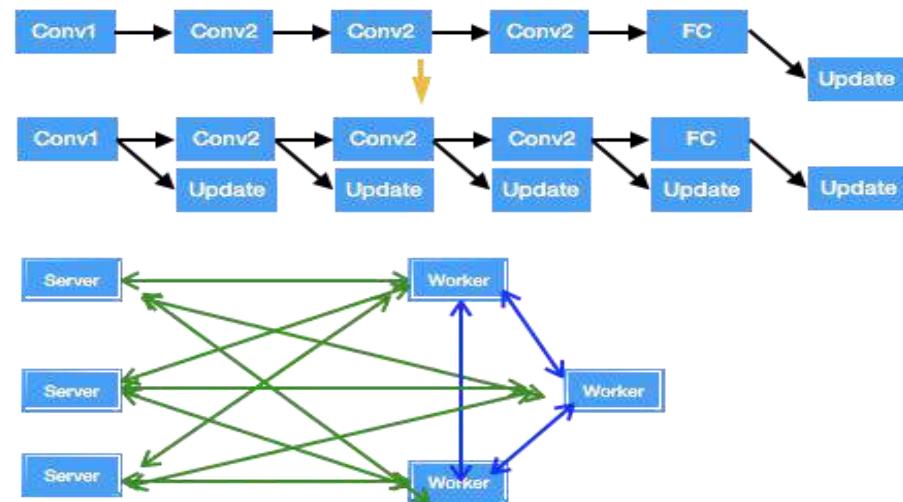
- 优化的优化

—

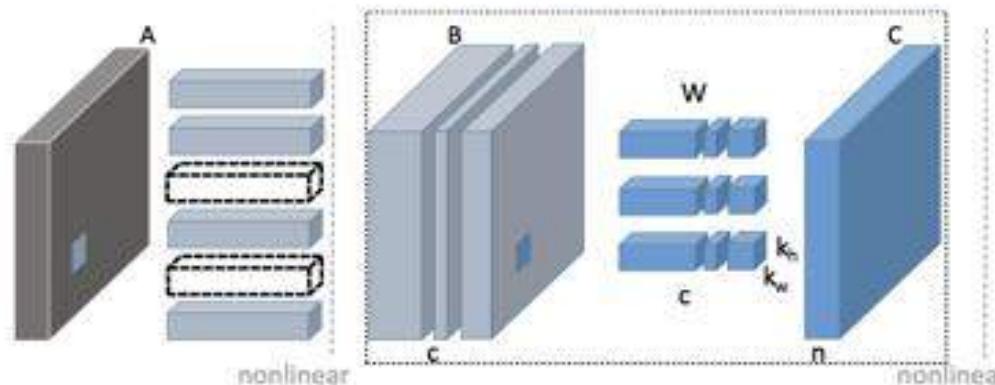


- 计算 通信重
 - 计算步骤 内存 负重
 - 计算步骤 络数 交换 新重

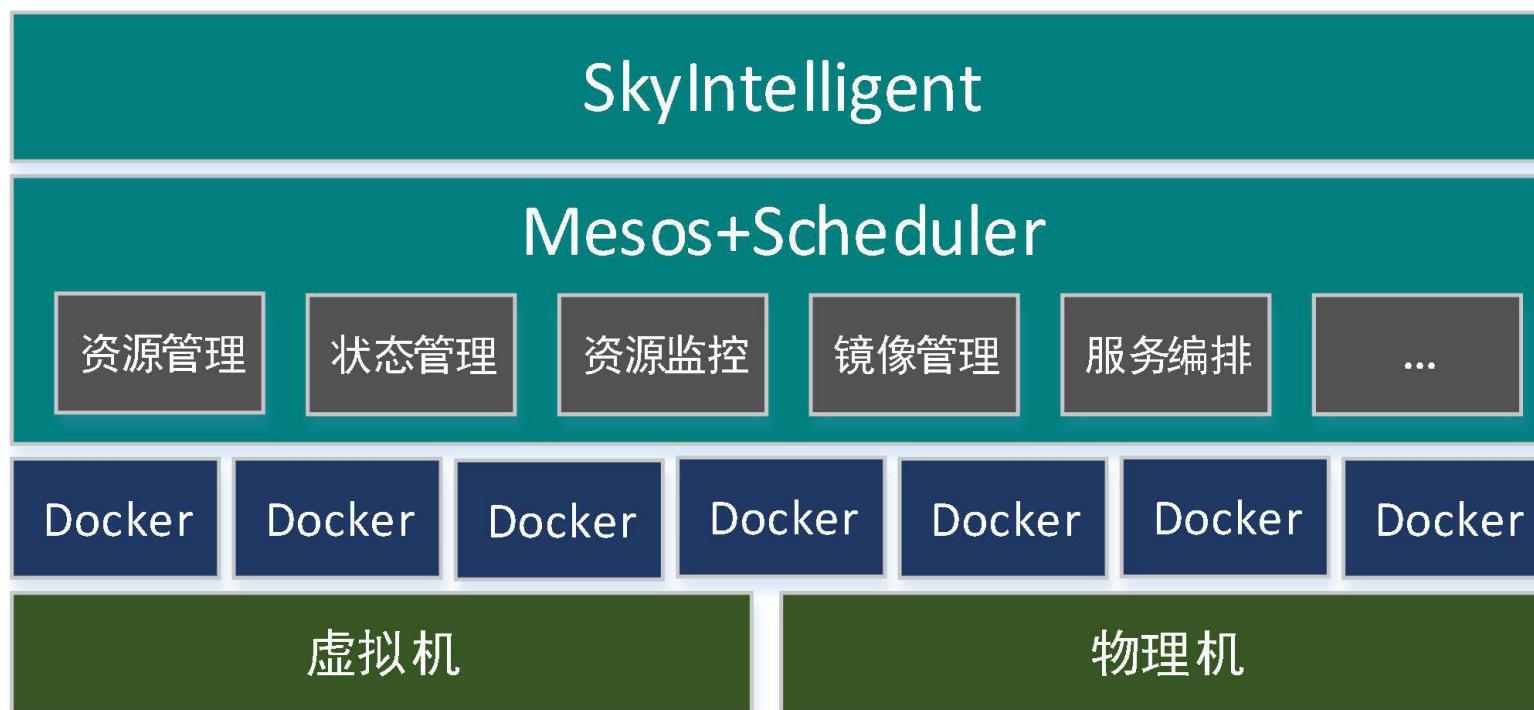
- 混合通信方式
 - 通过参数服务器传
 递小规模 度
 - 通过 交换大规模
 度
- 实现在 上的
 优化方法



- 通道 枝
 - 使用 来选择 枝目标
 - 使用最小二 来最小化重 误差
 - 在不 率的前提下可以 倍的推 理速度提
 - 同时 少 行时的内存 显存 用

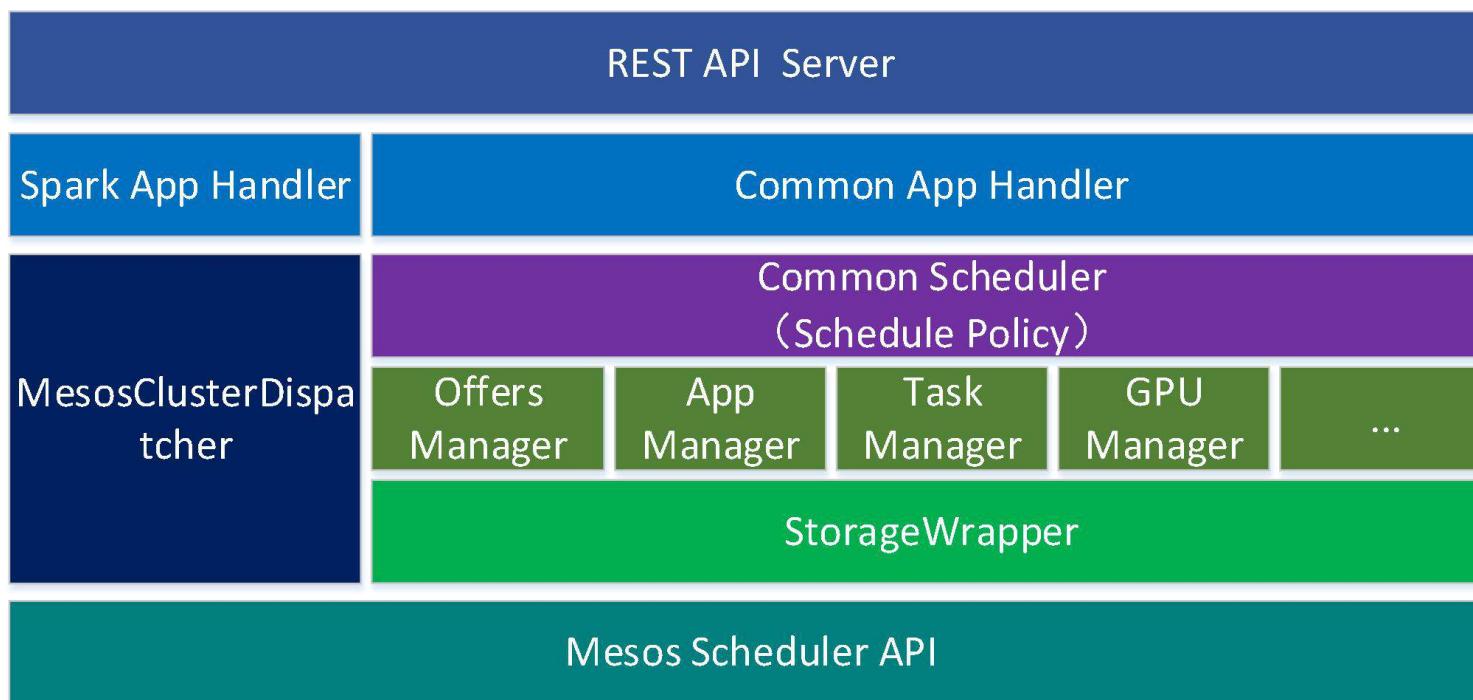


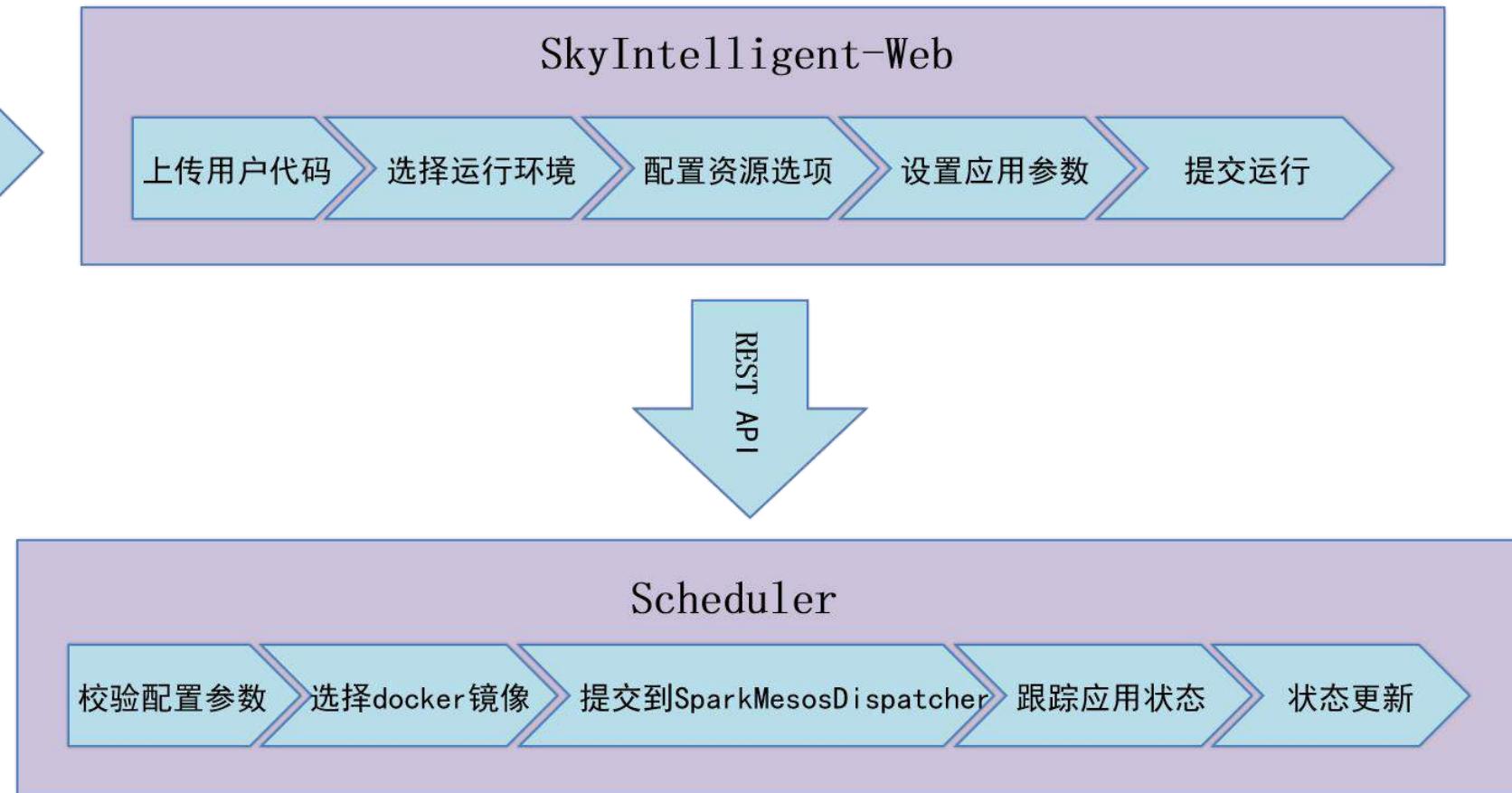
基 的容器化编 调度系统

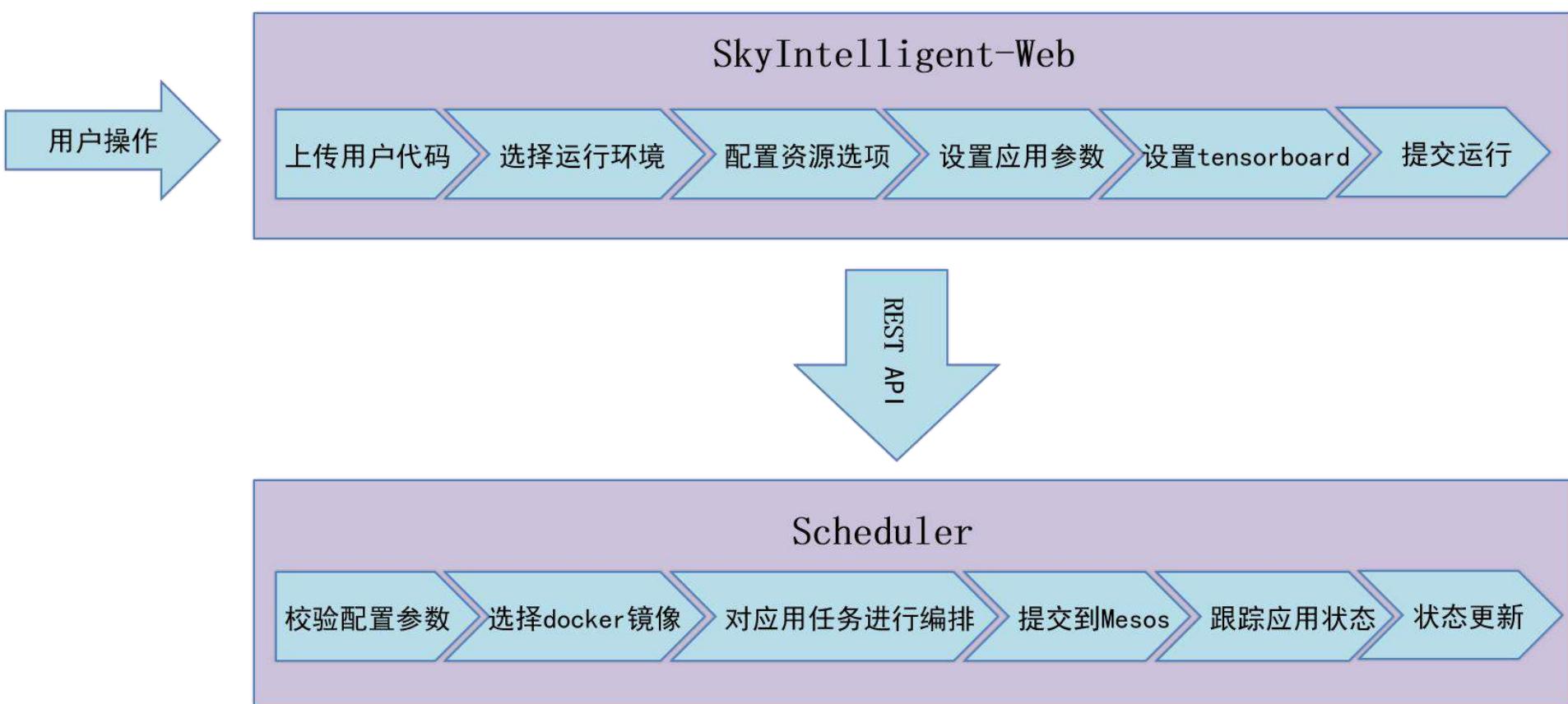


支持
深度学习

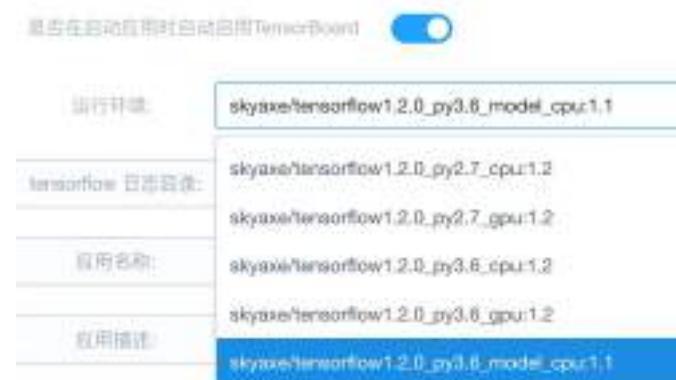
以
他较为流行的机器学习







- 行环境 开发环境的用户自定义
- 支持在线定制 保存镜像文件
- 支持用户上传镜像 配置服务



H2O FLOW Flow+ Cell+ Data+ Model+ Score+ Admin+

Untitled Flow

assist

Assistance

Routine	Description
ImportFiles	Import file(s) into H2O
getFrames	Get a list of frames in H2O
splitFrame	Split a frame into two or more
mergeFrames	Merge two frames into one

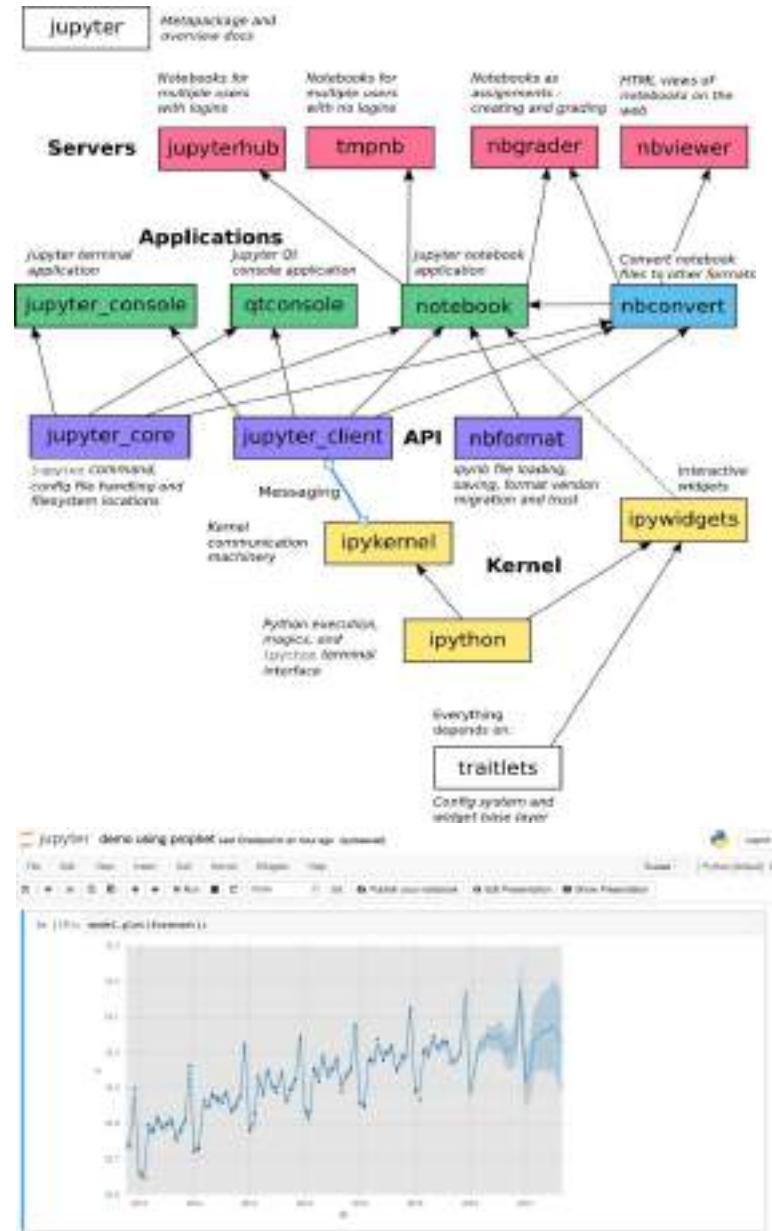
Deep Learning...
Deep Water...
Distributed Random Forest...
Gradient Boosting Machine...
Generalized Linear Modeling...
Generalized Low Rank Modeling...
K-means...
Naive Bayes...
Principal Components Analysis...
Stacked Ensemble...
Word2Vec...

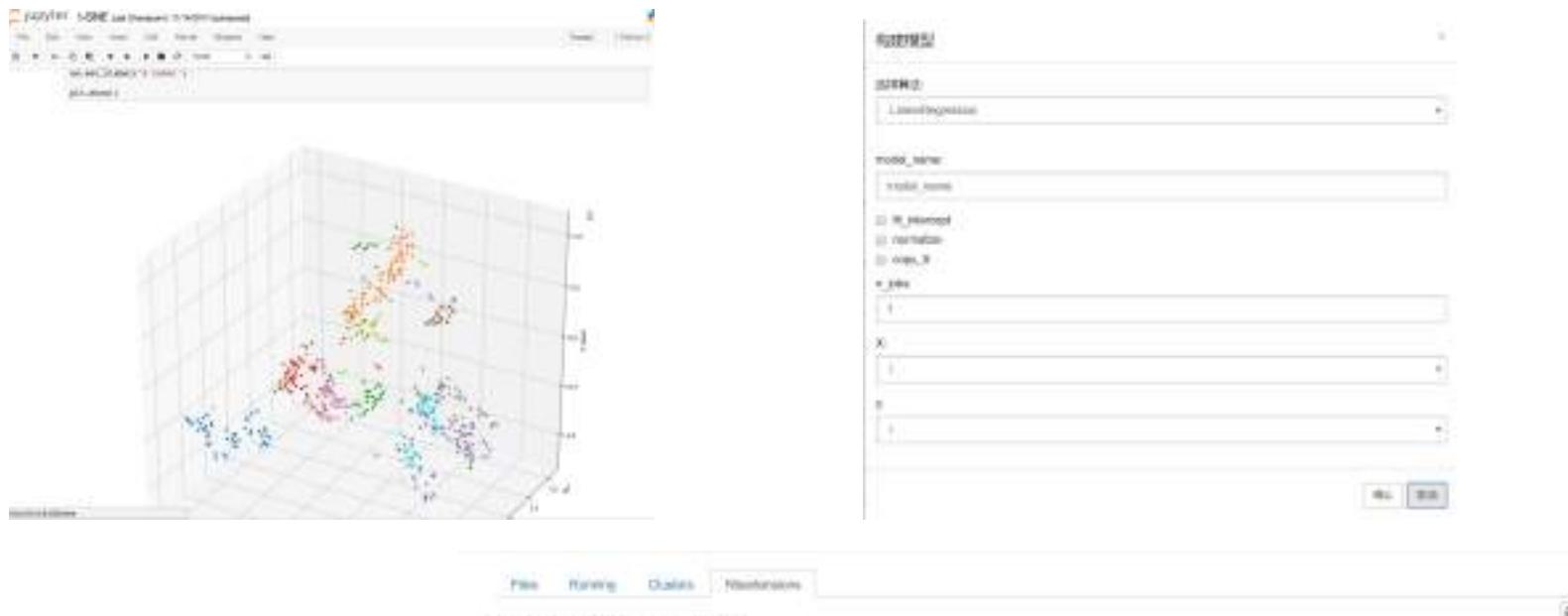
H2O H2O 运行中 打开工作空间 > 停止 详情

Nvidia_digits 停止中 详情



- 项目在开发开源软件开放标准用数十种编程语言的分布式计算服务
- 作为应用提供创建编辑包代码方程式图表以述性文字的文可用数清洗转换统计建模数可视化机器学习等方面
- 功能强大大量的使用者开发者资源共
- 可视化分布式分组件富
- 可扩展





Configurable nibextensions

✓ disable configuration for checkboxes without explicit compatibility layer may break your database environment, but can be useful to avoid schema evolution

Рівні: 1000000000, 100000000, 10000000

- Joiner's List/Checklist
 - AutoSaveTree
 - Code pretty
 - Collapsible Headings
 - Equations Auto Numbering
 - Export Embedded HTML
 - Hide Header
 - Nightlight
 - Keyboard shortcut editor
 - Navigation-Hotkeys
 - Notify
 - Rule
 - Select CodeBlocker Hotkeys
 - Snippets

[View Table of Contents \(2\)](#)

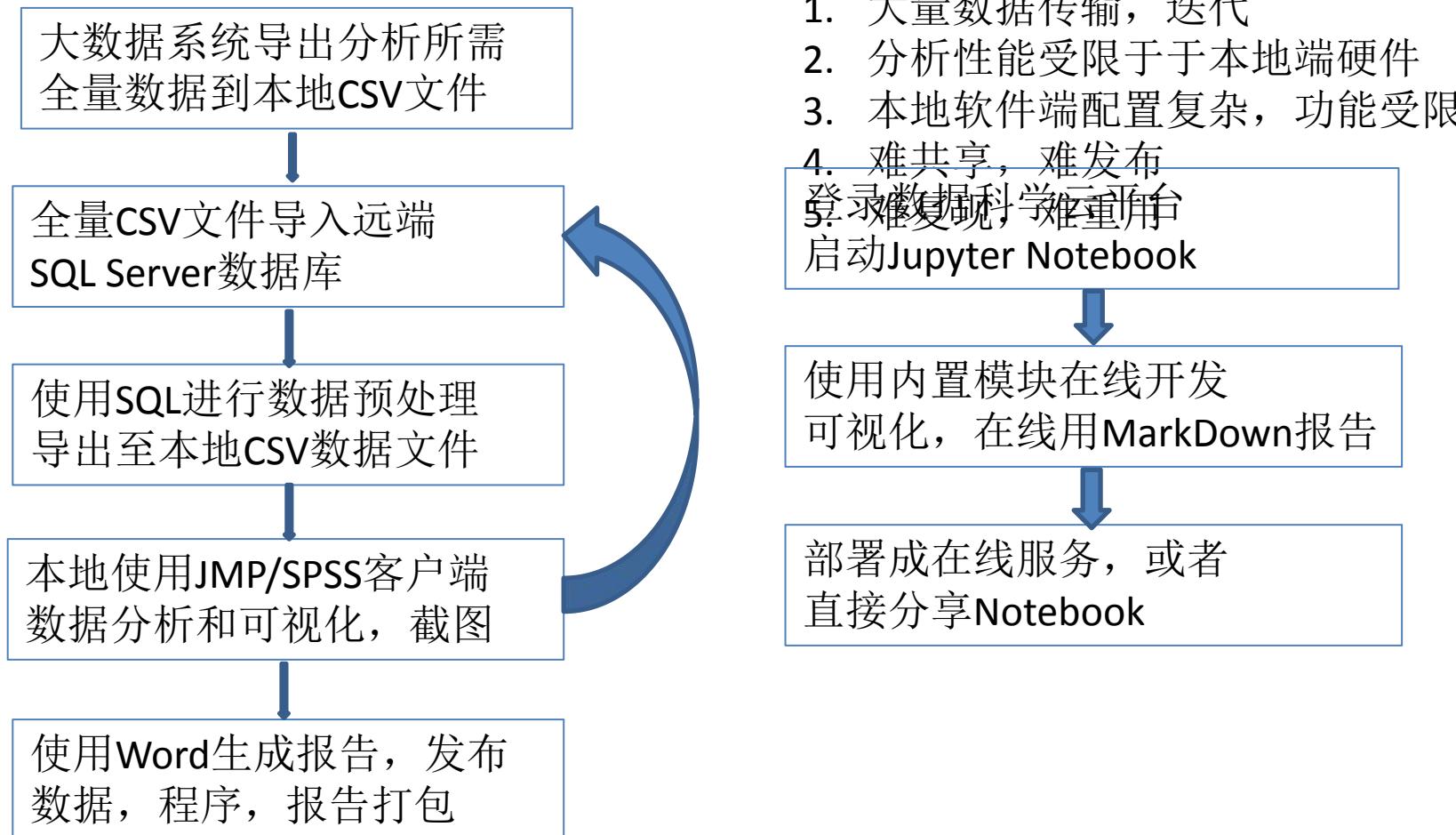
[View Variable Inspector](#)

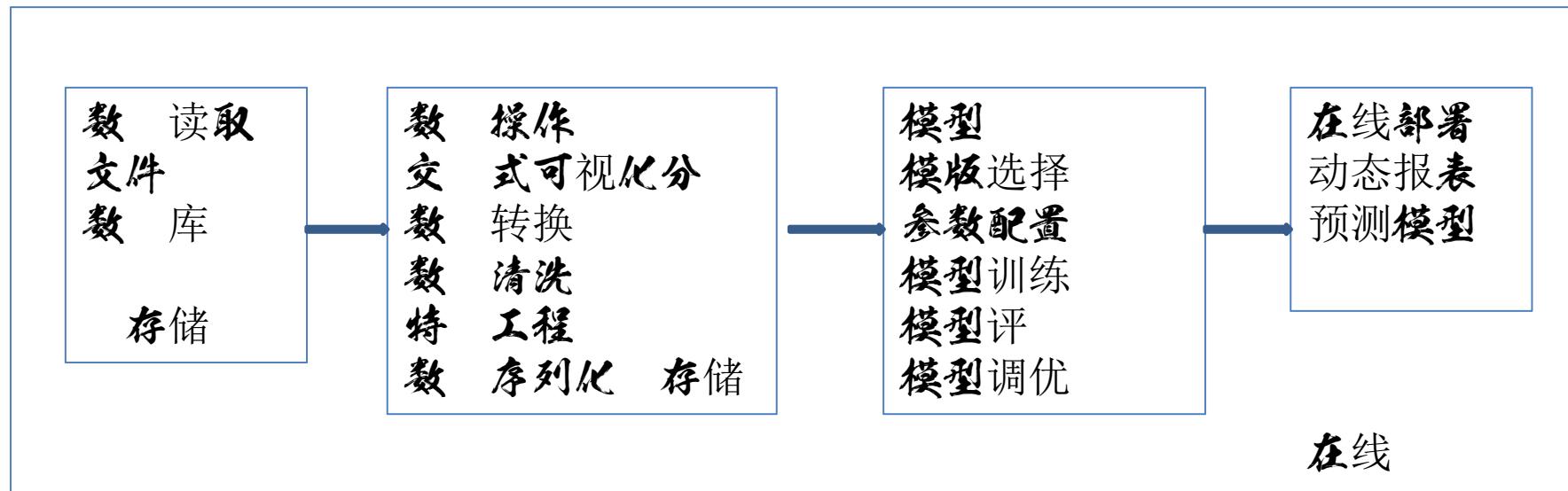
- [New Converter](#)
 - [Autoscale](#)
 - [Customizing](#)
 - [Compare Measurement History](#)
 - [Create Time](#)
 - [Pause](#)
 - [Hide Input](#)
 - [History](#)
 - [Launch QP Converter](#)
 - [Measurement Analysis](#)
 - [Print](#)
 - [Runnels](#)
 - [SHELL Syntax](#)
 - [Snippets Menu](#)
 - [Tabulator](#)
 - [Uzermode](#)

- [Add/Remove](#)
 - [Annotate/Highlight](#)
 - [Colorizing in Editor](#)
 - [Copy/Paste, reformatting, etc.](#)
 - [Exercise](#)
 - [Quiz](#)
 - [Write input all](#)
 - [Initialize cells](#)
 - [Load output](#)
 - [Modifications will reexecute](#)
 - [Python Monks](#)
 - [Scrapbook](#)
 - [Skip/Replace](#)
 - [Watchtower](#)
 - [Toggle all line numbers](#)

- Autoreload
 - Cache Font Size
 - CacheMore recent resources
 - Clear history
 - Download
 - Help panel
 - Highlight selected word
 - Print preview enhancement**
 - Move selected cells
 - Preferences
 - Hide sidebar
 - Show Smileyface**
 - Split Cells horizontally
 - Tree File

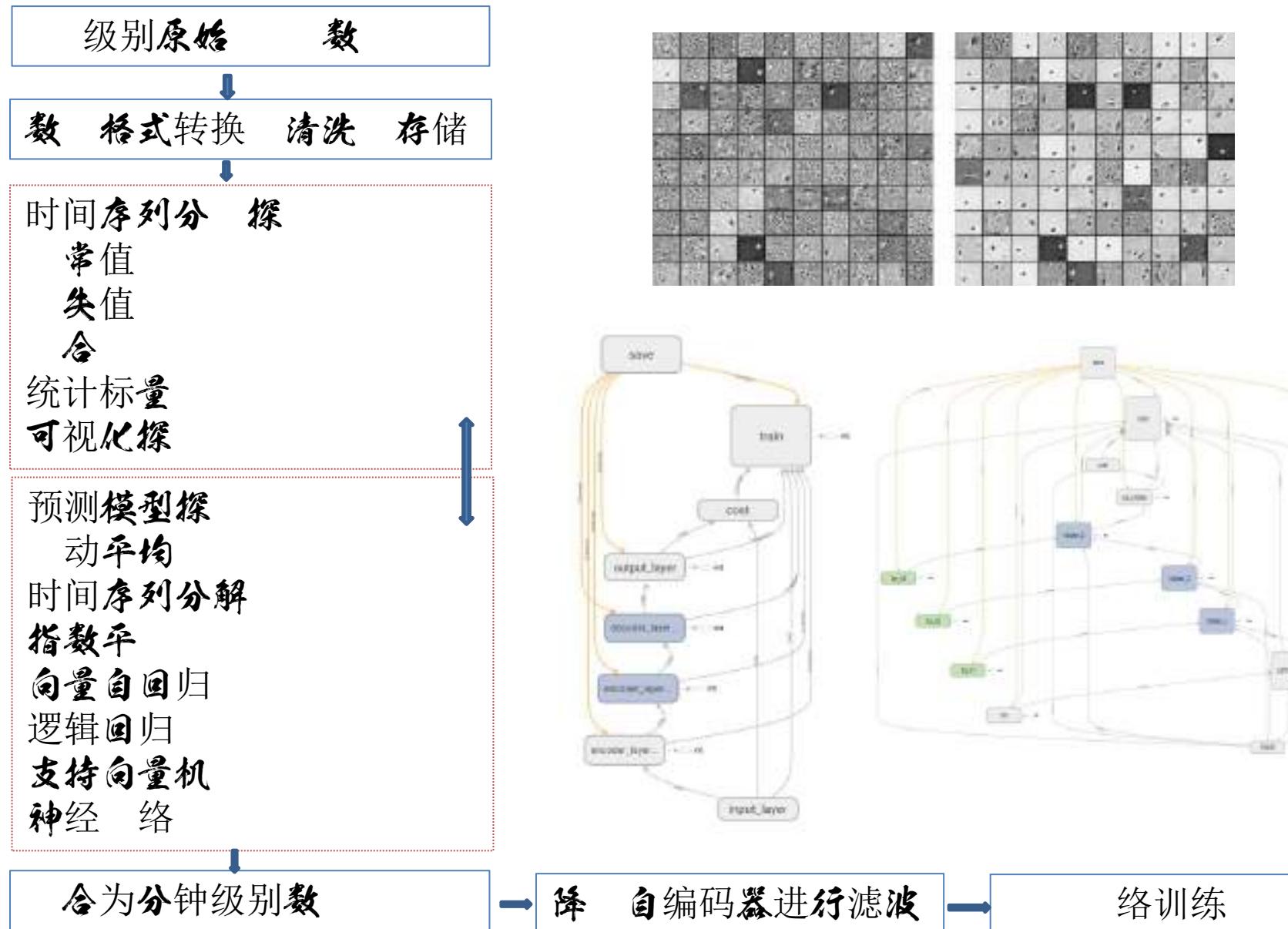
- 使用部署的模型 协议
- 使用的可视化库 应用 快速 部署 可视化 动态
视图 报表
- 通过 应用 来转换 部署 服务
- 所有部署使用可选择的 镜像 口映射 部署
均通过 来协调 调度





大数 量模型的 线训练
大规模计算的任务提交
线开发应用的 提交
线批量预测任务的提交

线



- 在线 模块 模版的终 用户的自由定制
- 公有 部署 支持
- 多针对时序型 实时数 的优化
- 多内置的模型 行业模版
- 多内置文 式可视化组件
-
-



谢谢！

Ling.Ni@sky-data.cn

BDTC 2017 中国大数据技术大会
Big Data Technology Conference 2017